Additional Table 6. Risk of bias analysis of the included studies - Downs and Black (1998) scale.

Nr.		1		2	3		4	5
Study*		**Abdelnaby and Nassar 2010**		**Altuğ et al. 1989**	**Barrett et al. 2010**		**Gökalp and Kurt 2005**	**Tuncer et al. 2009**
Reporting	*Is the hypothesis/aim/objective of the study clearly described?*	1		1	1		1	1
	Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1		1	1		1	1
	Are the characteristics of the patients included clearly described?	1		0	0		0	1
	Are the functional appliances used clearly described?	1		0	1		0	1
	Are the distributors of principal confounders in each group of subjects to be compared clearly described	1		0	1		1	1
	Are the main findings of the study clearly described?	1		1	1		1	1
	Does the study provide estimates of the random variability in the data for the main outcomes?	1		1	1		1	1
	Have all important adverse events that may be a consequence of functional appliances been reported?	1		0	0		1	1
	Have the characteristics of patients lost to follow-up been described?	0		0	0		0	0
	Have actual probability values been reported for the main outcomes except where the probability value is less than 0.001?	0		0	1		1	1
External validity	*Were the patients asked to participate in the study representative of the entire population from which they were recruited?*	0		1	0		0	0
	Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	0		1	0		0	0
	Were the staff, places and facilities where the patients were treated representative of the treatment the majority of patients receive?	0		1	1		1	0
Internal validity - bias	*Was an attempt made to blind study subjects for the intervention they had received?*	1		0	0		0	0
	Was an attempt made to blind those measuring the main outcome of the intervention?	0		0	1		0	0
	If any of the results of the study were based on "data dredging", was that made clear?	1		1	1		1	1
	In trials, do the analyses adjust for different lengths of follow-up of patients, or in case control studies, is the time and period between the intervention and outcome the same for cases and controls?	1		1	0		1	1
	Were the statistical tests used to assess the main outcomes appropriate?	1		1	1		1	1
	Was compliance with the extraoral appliance used reliable?	0		0	0		0	0
	Were the main outcome measures used accurate (valid and reliable)?	1		1	1		1	1
Internal validity – confounding (selection bias)	*Were the patients in different intervention groups recruited from the same population?*	1		1	0		0	1
	Were study subjects in different intervention groups recruited over the same period of time?	1		1	1		1	1
	Were study subjects randomized to intervention groups?	1		0	0		0	0
	Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	0		0	0		0	0
	Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	1		0	1		1	1
	Were losses of patients to follow-up taken into account?	0		0	0		0	0
		a	**b**		**a**	**b**		
Power	*Did the study (or the independent treated groups of the study where this applicable) have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?*	4	4	4	2	1	2	4
Sum		**21**	**21**	**17**	**16**	**15**	**16**	**20**

**Authors in alphabetical order; yes=1, no=0, unable to determine=0.*
Answers are scored 0 or 1, except for one item in the reporting domain, which is scored 0 to 2, and the single item on power, which is scored 0 to 5.
The correspondence between the sample sizes (N) and the power of the study, after applying G-power statistics, ranks as following: N<10 → power=0, 10<N<12 →power=1, 13<N<15 →power=2, 16<N<18 →power=3, 19<N<21 →power=4 and N>21 →power=5.